EXHIBIT 142

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

RICHARD KADREY, an individual; et al., v. META PLATFORMS, INC., a Delaware corporation; Defendant.	Case No. 3:23-cv-03417-VC

EXPERT REPORT OF PROF. MICHAEL SINKINSON, PH.D.

Signed in Evanston, Illinois on January 10, 2025

_____

Michael Sinkinson, Ph.D.

and audio books has been steadily increasing. According to the Association of American Publisher's annual report in 2023, digital and audio books made up 21 percent of total trade books (as defined below) sold, with e-book revenues up by 0.6 percent from the previous year.[122] Another 2023 survey found that 53 percent of American adults have listened to an audiobook, a 7 percent increase from the previous year, indicating an increasing demand for digital format books.[123]

50.   There are approximately 49,450 professional authors in the United States according to the U.S. Bureau of Labor Statistics' 2023 report.[124] This was down from an all-time high of 54,010 authors in 2022.[125] Numbers are not yet available for 2024. These statistics are a generous view of the total number of authors publishing books every year, as it also includes writers and authors in advertising, news and media.[126] There are also a number of authors whose primary profession is not as a writer, and they are therefore not accounted for in this data.[127]

51.   There are five major publishing houses, including Penguin Random House, HarperCollins, Macmillan, Simon & Schuster, and Hachette, which collectively account for roughly 80 percent of the U.S. market for trade books, while the remaining 20 percent of the market includes a variety of small-to-medium publishing houses.[128] As I will explain below in Section V.C, the term *trade book* in the book publishing industry generally refers to a book targeted toward a general audience and available through a variety of sales channels (as opposed to textbooks or reference books targeted towards educational, academic or professional markets). Additionally, authors have the ability to self-publish their books through services such as Amazon's Kindle Direct Publishing (KDP), launched in 2007.[129] Over the last decade, the self-publishing industry has experienced massive growth, as evidenced by the significant increase in the number of ISBNs issued for self-published works. In 2011, the total number of self-published ISBNs issued stood

---

[122] Porter Anderson, "AAP StatShot: In 2023, US Revenues Were $12.6 Billion," Publishing Perspectives, 26 Mar 2024, available at https://publishingperspectives.com/2024/03/aap-statshot-in-2023-us-revenues-were-12-6-billion/, accessed on 29 Sep 2024.

[123] Porter Anderson, "APA Report: US Sees Double-Digit Audiobook Sales Growth for 11th Year," Publishing Perspectives, 1 Jun 2023, available at https://publishingperspectives.com/2023/06/apa-report-audio-sales-in-double-digits-for-11th-year/, accessed on 29 Sep 2024.

[124] Amy Watson, "Number of writers and authors in the U.S. 2011-2023," Statista, 12 Apr 2024, available at https://www.statista.com/statistics/572476/number-writers-authors-usa/#statisticContainer, accessed on 29 Sep 2024.

[125] Amy Watson, "Number of writers and authors in the U.S. 2011-2023," Statista, 12 Apr 2024, available at https://www.statista.com/statistics/572476/number-writers-authors-usa/#statisticContainer, accessed on 29 Sep 2024.

[126] "Occupational Employment and Wage Statistics, May 2023," U.S. Bureau of Labor Statistics, available at https://www.bls.gov/oes/current/oes273043.htm, accessed on 29 Sep 2024.

[127] "Occupational Employment and Wage Statistics, May 2023," U.S. Bureau of Labor Statistics, available at https://www.bls.gov/oes/current/oes273043.htm, accessed on 29 Sep 2024.

[128] Magda Wojcik, "5 biggest book publishing companies: A guide for authors," MW Editing, 12 Feb 2024, available at https://www.mwediting.com/biggest-book-publishing-companies/, accessed on 29 Sep 2024.

[129] "How Amazon's self-publishing service has empowered authors and diversified publishing over the last 15 years," Amazon, 17 Nov 2022, available at https://www.aboutamazon.com/news/books-and-authors/amazon-self-publishing-empowers-authors, accessed on 29 Sep 2024.

at 526,907.[130] By 2021, this figure had increased to 2,298,004, a 336 percent increase over ten years.[131] Notably, the roughly 2.3 million titles self-published in the U.S. in 2021, is also significantly higher than the 500,000 to 1 million books that were published through established publishers in the U.S. market.[132] Overall, there are millions of new books published every year through traditional publishers or self-publishing platforms.

2) Copyright Laws Protect an Author's Original Work to Incentivize the Production of Novel Creative Works

52. I am not a lawyer, so for purposes of my analysis, I rely on Counsel to inform my understanding of the relevant provisions of the copyright laws. In this section, I will provide an overview of that understanding, and further discussion of the economic underpinnings and implications of those laws.

53. Article I, Section 8, Clause 8 or the U.S. Constitution secures to Congress the "Power … [t]o promote the Progress of Science and useful Arts" through the issuance of patents and copyrights. Intellectual property laws, generally speaking, seek to incentivize the production and publication of novel intellectual works, such as inventions and creative works, for the benefit of the public.[133] Copyright protection enables authors to monetize and safeguard their creative works by giving them recourse against unauthorized use.[134]

54. Copyright protection is granted to an original work of authorship created by a human author that exhibits a minimum threshold of originality or creativity.[135] Copyright protection does not extend to ideas, processes, concepts, or discoveries within a work, although it may still protect their unique expression.[136]

---

[130] Jim Milliot, "Self-Publishing Is Thriving, According to Bowker Report," Publishers Weekly, 17 Feb 2023, available at https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/91574-self-publishing-is-thriving-according-to-bowker-report.html, accessed on 4 Jan 2025.

[131] Jim Milliot, "Self-Publishing Is Thriving, According to Bowker Report," Publishers Weekly, 17 Feb 2023, available at https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/91574-self-publishing-is-thriving-according-to-bowker-report.html, accessed on 4 Jan 2025.

[132] Rob Errera, "How Many Books Are Published Each Year? [2024 Statistics]," Toner Buzz, 18 Aug 2023, available at https://www.tonerbuzz.com/blog/how-many-books-are-published-each-year/, accessed on 30 Sep 2024.

[133] Congress is empowered to "Promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." *See* Article I, section 8, clause 8 of the United States Constitution, available at https://www.archives.gov/founding-docs/constitution-transcript, accessed on 30 Sep 2024; *see also* "Intellectual Property," Cornell Law School, Jun 2023, available at https://www.law.cornell.edu/wex/intellectual_property, accessed 30 Sep 2024.

[134] "Intellectual Property," Cornell Law School, Jun 2023, available at https://www.law.cornell.edu/wex/intellectual_property, accessed on 30 Sep 2024.

[135] The U.S. Supreme Court has emphasized that a work must possess a "spark" and "modicum" of creativity to qualify. In Feist Publications v. Rural Telephone Service, 499 U.S. 340 (1991), the Supreme Court held that a work must have "some minimal degree of creativity." *See* "Copyright," Cornell Law School: Legal Information Institute, available at https://www.law.cornell.edu/wex/copyright, accessed on 30 Sep 2024.

[136] This is codified in 17 U.S.C. § 102, which states: "In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work." Feist Pub's, Inc. v. Rural Telephone Svc. Co., 499 U. S. 340, 359 (1991) ("[F]acts contained in existing works may be freely copied."); *see* "Copyright," Cornell Law School: Legal Information Institute, available at https://www.law.cornell.edu/wex/copyright, accessed on 30 Sep 2024.

market to Llama. To do this, in Section VI.A, I provide an empirical analysis of whether there is any indication that the release of Llama has led to any actual substitution away from original works. In Section VI.B, I consider factors related to substitutability of the original works. I examine the potentially positive effects of the release of Llama on the original market in Section VI.C. While I find no indication of substitution or substitutability in the original market, an economic analysis related to the capabilities of Llama indicates that the release of Llama likely has a positive effect on authors' ability to market their books in their original market. Overall, I do not find any empirical or theoretical evidence of a negative effect of Llama on the original market for trade books.

A. Book Sales Are Not Negatively Affected by the Release of Llama 3, as Evidenced by My Empirical Analysis of Book Sales

1) The Effect of the Release of Llama 3 on Sales of Plaintiffs' Books Is an Appropriate Scope for My Empirical Analysis

93.   In this section, I examine the empirical evidence on whether the release of Llama 3 affected the sales of books allegedly used as training data for Llama. As explained above, this analysis is relevant to Factor 4 of fair use, that is, "the effect of the use upon the potential market for or value of the copyrighted work."

94.   There are various potential mechanisms through which the release of Llama may, in principle, have affected the sales of books allegedly used as AI training data. First, the release of Llama could theoretically have provided a means to consume the content of books used as AI training data as an alternative to purchasing the books themselves. Second, the release of Llama could hypothetically have provided consumers with a complementary (rather than substitute) use that boosts demand and purchases for books used as AI training data, as discussed in Section V.D. Third, the release of Llama could have provided consumers with a complementary (rather than substitute) use, such as criticism and commentary, that dampens demand for such books; this would not, however, be a cognizable harm under the Copyright Act, as discussed in Section V.B.[190]

95.   To estimate the potential effects of the release of Llama 3 on the sales of books allegedly used as training data for Llama, I rely on an econometric estimation methodology called *first differences*, which is commonly used in empirical settings to study changes in outcomes over time.[191] My methodology is supported by *robustness checks* which demonstrate that the conclusions of my analysis do not depend on any particular set of assumptions.

96.   My analysis focuses on the effects of the release of Llama 3, rather than the earlier releases of Llama 1 on February 24, 2023 or Llama 2 on July 18, 2023. Meta integrated Llama into its products and services at

---

[190] As noted above, the Supreme Court has held that the latter possibility—that a secondary use such as criticism or parody "kills demand for the original"—"does not produce a harm cognizable under the Copyright Act." *See Campbell* v *Acuff-Rose Music*, 510 U.S. 569, 10 591–92 (1994).

[191] "The First Difference Estimator," Tilburg Science Hub, available at https://tilburgsciencehub.com/topics/analyze/causal-inference/panel-data/first-difference/, accessed on 18 Dec 2024.

Table 6. First difference regression results with alternate 'After' durations

Variables	1 Week	2 Weeks	3 Weeks	4 Weeks
[A]	[B]	[C]	[D]	[E]
[1] After Release	-0.003**	-0.001	-0.001	-0.002***
[2]	[-0.006, 0.000]	[-0.003, 0.001]	[-0.002, 0.001]	[-0.003, -0.001]
[3] Intercept	0.003***	0.003***	0.003***	0.003***
[4]	[0.003, 0.003]	[0.003, 0.003]	[0.003, 0.003]	[0.003, 0.003]
[5] Total Obs.	8,698,339	8,749,390	8,800,027	8,851,032

	5 Weeks	6 Weeks	7 Weeks	8 Weeks
	[F]	[G]	[H]	[I]
[1] After Release	-0.003***	-0.003***	-0.003***	-0.003***
[2]	[-0.004, -0.002]	[-0.004, -0.003]	[-0.003, -0.002]	[-0.004, -0.002]
[3] Intercept	0.003***	0.003***	0.003***	0.003***
[4]	[0.003, 0.003]	[0.003, 0.003]	[0.003, 0.003]	[0.003, 0.003]
[5] Total Obs.	8,901,296	8,951,884	9,002,522	9,053,037

Sources and Notes:
[A]: Regression Output Variables.
[B] … [I]: Duration of 'After' period starting the week of Llama 3's release (April 18, 2024).
[1] & [6]: Coefficient estimates for 'After' in first difference regression equation.
[2] & [7]: Confidence interval for [1] 'After'.
[3] & [8] Intercept estimate in first difference regression equation.
[4] & [9] Confidence interval for [3] intercept.
[5] & [10] Count of total observations in regression estimation.
*, **, and *** denote statistical significance at the 0.1, 0.05, and 0.01 level.

281. The results provide me with a pair of estimates pertaining to value and precision of β. First is a *point estimate*, which is an estimate of β and is shown in row [1], and second is a *standard error*, which is an indication of how precise the point estimate is. With these two estimates, I can obtain a *confidence interval* for β, which is shown in row [2]. A confidence interval is an interval around the point estimate that provides, for a desired confidence level, the likelihood that the true value of β lies within that interval.

282. In the results in Table 6, I provide the point estimate of β and the associated 95 percent confidence interval that indicates a 95 percent confidence that β lies within this interval. [471]

283. Table 6 indicates that in the first three weeks following Llama 3's release (Column [E]), there is no detectable change[472] in relative sales ranks for books in the sample relative to all other books tracked by

---

[471] According to statistical properties, the 95 percent confidence interval of a point estimate is approximately 1.96 times the standard error above and below the point estimate. This means that for a hypothetical point estimate of 10 and a standard error of 2, the 95 percent confidence interval is 10 ± (1.96 x 2), or the interval (6.08, 13.92).
[472] Up to a precision of 3 decimal points.

Keepa. In the first eight weeks after Llama 3's release (Column [I]), there was an approximately an 0.3 percent decline in the numerical value of the sales ranks (i.e., an improvement in ranks) for the books in sample. This corresponds to an improvement of approximately 3 spots for a book that is previously ranked 1000, or a rise in 3000 spots for a book that is previously ranked 1,000,000.

284. Table 6 also indicates the estimated value of the intercept, α, in row [3] and the associated 95 percent confidence interval of α in row [4]. The estimate of 0.003, which corresponds to approximately 0.3 percent, indicates that books typically exhibit a trend of a 0.3 percent decline in sales rank from week to week. This value, which is statistically significant and precisely estimated, is consistent with the earlier Figure 12 illustrating the typical book's sales rank trend.

285. Table 6 indicates no substantive change in sales rank for books in the Books3 dataset relative to all other books tracked by Keepa, and does not detect a detrimental impact.

4) Placebo Comparison to Validate the Negligibility of the Treatment Effect

286. To assess the robustness of my empirical findings, I employ a "placebo test" using a placebo sample. In statistics, a placebo sample is a sample of the data which is not exposed to the treatment of interest.[473]

287. I examine whether there is a "response" in sales rank following weeks when Llama 3 was not released yet. Specifically, I estimate the FD regression for each week between June 2021 and April 2024, which provides me a total of 151 coefficient estimates. One of these estimates – the last one – corresponds to the release of Llama 3, which is compared against the distribution of earlier placebo estimates to detect whether it is significantly different. I repeat this exercise for the different treatment durations considered in my main analysis (1-8 weeks). Any estimations where the week duration is larger than 1 week and overlaps with the week of the release of Llama are dropped to avoid potentially picking up a response to the release of Llama.

288. In Table 7 I provide summary statistics on the eight distributions of placebo coefficient estimates. In Column [B], I show the percent of estimates that are larger in value than the estimate corresponding to Llama 3's release, and in Column [C] I show the percent of estimates that are larger in magnitude than the estimate corresponding to Llama 3's release, each separately for the eight placebo distributions corresponding to the duration of the "After" period. They summarily indicate that a substantial share of coefficients is numerically larger or further from 0 than the estimate corresponding to Llama 3's release, indicating that during the few months shortly following Llama 3's release, books in the sample exhibited typical sales rank patterns. The remaining columns in Table 7 provide the mean, median, 5th percentile and 95th percentile of each placebo distribution, and highlight the narrow bound for the typical change in sales rank patterns.

---

[473] Ting Ye, Shuxiao Chen and Bo Zhang, "The Role of Placebo Samples in Observational Studies," ArXiv, 5 Jul 2022, available at https://arxiv.org/pdf/2205.10761.

Table 7. Summary Statistics of Placebo Distributions

Duration	Percent Larger Than Llama 3 Estimate	Percent Further From 0	Mean	Median	5th Percentile	95th Percentile
[A]	[B]	[C]	[D]	[E]	[F]	[G]
[1] 1 Week	65%	64.00%	0.000	-0.001	-0.016	0.017
[2] 2 Weeks	49%	83.89%	0.000	-0.001	-0.012	0.011
[3] 3 Weeks	45%	88.51%	0.000	-0.001	-0.008	0.009
[4] 4 Weeks	59%	62.59%	0.000	-0.001	-0.006	0.009
[5] 5 Weeks	83%	34.93%	0.000	-0.001	-0.007	0.008
[6] 6 Weeks	83%	34.48%	0.000	-0.001	-0.006	0.008
[7] 7 Weeks	82%	37.50%	0.000	-0.001	-0.005	0.007
[8] 8 Weeks	83%	32.17%	0.000	-0.001	-0.005	0.007

Sources and Notes: Analysis of Keepa sales rank data.
[A]: Duration of "After" period in regression equation.
[B]: The percent of placebo coefficient estimates that are numerically larger (more positive) than the coefficient estimate corresponding to Llama 3's release
[C]: The percent of placebo coefficient estimates that are further from zero (larger in absolute value) than the coefficient estimate corresponding to Llama 3's release
[D]: The mean of the placebo distribution
[E]: The median of the placebo distribution
[F]: The 5th percentile of the placebo distribution
[G]: The 95th percentile of the placebo distribution.

289. Consistent with the previous results, Table 7 indicates that the changes in sales rank pattern in the weeks following Llama 3's release are highly typical relative to sales rank patterns in the preceding three years before Llama 3's release. These results indicate that Llama 3's release cannot be associated with a worsening of sales ranks among books that were used for training Llama 3.